a)

DUPLEX SYSTEM FOR MAGNETIC DISK CONTROLLER

Patent Number: JP7160432
Publication date: 1995-06-23

Inventor(s): ISHII TAKASHI
Applicant(s): TOSHIBA CORP

Requested Patent: ☐ JP7160432
Application Number: JP19930308995 19931209
Priority Number(s):
IPC Classification: G06F3/06; G06F11/16; G06F12/08
EC Classification:
Equivalents: JP3122295B2

Abstract

PURPOSE:To construct the duplex system for the magnetic disk controller which has an environment wherein the high reliability of a magnetic disk device having a write-back type disk cache is easily realized by copying cache data of a controller of one system to the other system at normal time. CONSTITUTION:The system is provided with a data transfer path 3 which copies the contents of the cache data and control information on the storage destination address, etc., of the data from the controller 10 that normally responds to a command from a host device to the controller 20 normally in a stand-by state, and, the same contents are made normally present in the caches 2 and 2 of the controllers of both the systems, but if the controller 10 gets out of order, the cache data are written to the magnetic disk device 5 from the controller 20.

a)

HEI 7-160432

[DETAILED DESCRIPTION OF THE INVENTION]

[0001]

[Industrial Application Field] The present invention relates to a duplex system for magnetic disk controller suitable for use in a write-back type disk cache system in which a plurality of magnetic disk controllers share a magnetic disk device and the magnetic disk controllers receive data produced by a host device to return a status to the effect of the completion of processing to the host device at that time and then write data in the magnetic disk device.

[0002]

[Description of Related Art] A magnetic disk device has frequently been used as a large-capacity external file unit not only in the field of large scale computer but also in the field of distributed processing computer, and even recently in the field of personal computer, and it has been toward the size reduction and densification year by year.

[0003] For securing the reliability of such a type of system, a system has been constructed which implements the duplex of magnetic disk controller for controlling a magnetic disk device. So far, such a magnetic disk controller duplexing system has employed a method in which an access path to the magnetic disk device is allocated to both the magnetic disk controllers to, in a case in which one magnetic disk controller falls into a trouble,

maintain the access from the other magnetic disk controller to the magnetic disk device.

[0004] The recent magnetic disk controller has more frequently employed a write-back type disk cache in which the magnetic disk controller first receives writing data from a host (host side device) and informs the host of a status to the effect of the completion of processing at that time and then carries out the data writing in the magnetic disk device. This disk cache forms an essential support item for fast access to the magnetic disk device.

[0005] In the case of the employment of the aforesaid write-back type disk cache, after the occurrence of a trouble of the magnetic disk controller, there is a need to write in the magnetic disk device the data left in a cache memory. Accordingly, in a conventional duplex system in which prepared simply are two magnetic disk controllers each having a cache memory, it is impossible to write end writing data in the magnetic disk device.

[0006] For this reason, in order to avoid this disadvantage, there has been known a system in which the cache memory is separated from the magnetic disk controllers and is made to have access paths from both the magnetic disk controllers. However, in the case of the employment of such a system, there is a need to make a connection of a special cache device even in the case of no duplex for the magnetic disk controller, which leads to a complicated hardware structure of the system and, hence, an increase in cost.

[0007]

[Problems to be Resolved by the Invention] In the case of the employment of the write-back type disk cache mentioned above, there is a need to write in the magnetic disk device the data left in the cache memory after the occurrence of a trouble of the magnetic disk controller and, hence, in the case of the conventional duplex system simply using two magnetic disk controllers each having a cache memory, impossibility is encountered in writing end writing data in the magnetic disk device. Therefore, it has been considered to employ a system in which the cache memory is separated from the magnetic disk controller and has an access path from both the magnetic disk controllers. However, even in the case of no duplex for the magnetic disk controller, there is a need to make a connection of a special cache device, which leads to a complicated hardware structure of the system and, hence, an increase in cost.

[0008] The present invention has been developed in consideration of the above-mentioned situations, and it is an object of the invention to provide a duplex system for the magnetic disk controller which has an environment in which a high reliability of a magnetic disk device having a write-back type disk cache is easily realized by copying cache data of one magnetic disk controller to the other magnetic disk controller at normal time.

[0009]

[Means for Resolving the Problems] In a disk cache system in which a plurality of magnetic disk controllers share a magnetic disk device and the magnetic disk controllers receive data produced by a host device and return a status

to the effect of the completion of processing to the host device at that time and then write data with respect to the magnetic disk device, the present invention is characterized in that a path for making communications between both the magnetic disk controllers is provided as an access path to the shared magnetic disk device and cache data of one magnetic disk controller is copied through the path into the other cache and both the magnetic disk controllers recognize the data, and data writing is made with respect to the magnetic disk device through the magnetic disk controller designated by the host device. Moreover, it is characterized in that the magnetic disk controller which is in operation always copies the cache data and sends it to the magnetic disk controller which is in a stand-by condition and, at the occurrence of a trouble of the magnetic disk controller which is in operation, the cache data is written in the magnetic disk device through the magnetic disk controller which is in the stand-by condition.

a)

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号

特開平7－160432

(43)公開日　平成７年(1995)６月23日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	３０４　Ｅ			
11/16	３１０　Ｆ			
12/08	３２０	7608－5Ｂ		

審査請求　未請求　請求項の数３　ＯＬ　（全　５　頁）

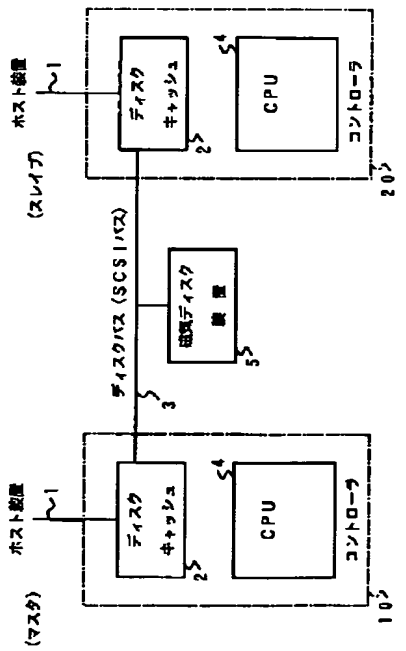
(21)出願番号　特願平5－308995

(22)出願日　平成５年(1993)12月９日

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者　石井　隆
東京都青梅市末広町２丁目９番地　株式会社東芝青梅工場内
(74)代理人　弁理士　鈴江　武彦

(54)【発明の名称】　磁気ディスクコントローラの二重化方式

(57)【要約】
【目的】この発明は、通常時に片系コントローラのキャッシュデータを他系に複製しておくことにより、ライトバック方式のディスクキャッシュを有する磁気ディスク装置の高信頼性を容易に実現する環境を持った磁気ディスクコントローラの二重化方式を構築することを主な特徴とする。
【構成】通常時にホスト装置からのコマンドに応答するコントローラ１０から、そのコントローラが故障した時に稼働するため、通常、スタンバイ状態にあるコントローラ２０へキャッシュデータの内容とデータの格納先アドレス等の制御情報を複写するデータ転送経路３を設け、通常時は両系のコントローラのキャッシュ２，２に同一の内容が存在するようにして稼働させ、コントローラ１０が故障時はコントローラ２０からキャッシュデータを磁気ディスク装置５へ書き込むことを特徴とする。



1

【特許請求の範囲】

【請求項１】　キャッシュをもつ複数の磁気ディスクコントローラが磁気ディスク装置を共有し、上記磁気ディスクコントローラがホスト装置より生成されるデータを受信すると、ホスト装置に対しその時点で処理終了の旨のステータスを返し、その後に磁気ディスク装置にデータの書き込みを行なうディスクキャッシュシステムに於いて、上記磁気ディスクコントローラ相互の間で通信を行うパスを、共有磁気ディスク装置に対するアクセスパスとして備え、このパスを介して一方の磁気ディスクコントローラが持つキャッシュデータを他方のキャッシュに複製し、双方の磁気ディスクコントローラにてデータの認知を行ない、ホスト装置によって指示される磁気ディスクコントローラ経由で磁気ディスク装置に対してデータの書き込みを行うことを特徴とする磁気ディスクコントローラの二重化方式。

【請求項２】　稼働状態にある磁気ディスクコントローラが、待機状態にある磁気ディスクコントローラに対し常にキャッシュデータを複製して送付し、故障時に、待機状態にある磁気ディスクコントローラを介してキャッシュデータを磁気ディスク装置に対し書き込むことを特徴とする請求項１記載の磁気ディスクコントローラの二重化方式。

【請求項３】　複数の磁気ディスクコントローラは、それぞれＣＰＵを内蔵し、このＣＰＵは、他方の磁気ディスクコントローラに対して書き込みアドレスを含むコマンドを発行し、ホスト装置から得られるデータを自系のキャッシュへ書き込むと同時にアクセスパスを介して他系のキャッシュへも書き込み、ホスト装置に対してステイタスを返した後、他方の磁気ディスクコントローラからコマンドを受信し、他方の磁気ディスクコントローラが磁気ディスク装置に対し書き込んだデータのディレクトリ情報を得、自系のディレクトリィ更新に備えることを特徴とする請求項１記載の磁気ディスクコントローラの二重化方式。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】この発明は、磁気ディスク装置を複数の磁気ディスクコントローラで共有し、ホスト装置によって生成されるデータを上記磁気ディスクコントローラが受信し、ホスト装置に対してその時点で処理終了の旨のステータスを返し、その後に磁気ディスク装置に対するデータの書き込みを行うライトバック方式を採るディスクキャッシュシステムに用いて好適な磁気ディスクコントローラの二重化方式に関する。

【０００２】

【従来の技術】磁気ディスク装置は大型コンピュータのみならず、分散処理コンピュータ、最近ではパーソナルコンピュータの分野に於いても大容量外部ファイル装置として多用され、年々、小型化、高密度化されてきている。

2

【０００３】この種システムの信頼性を確保するために、磁気ディスク装置のコントロールを行う磁気ディスクコントローラを二重化したシステムが構築されている。従来、このような磁気ディスクコントローラを二重化したシステムに於いては、磁気ディスク装置に対するアクセスパスを双方の磁気ディスクコントローラに対して持たせ、片系の磁気ディスクコントローラが故障した場合に、他系の磁気ディスクコントローラから磁気ディスク装置へのアクセスが維持できる方法が採られていた。

【０００４】最近の磁気ディスクコントローラでは、ホスト（ホスト側の装置）からの書き込みデータを磁気ディスクコントローラが先ず受取り、ホストに対しその時点で処理終了の旨のステータスを報告し、その後に磁気ディスク装置へのデータ書き込みを行うライトバック方式のディスクキャッシュが用いられることが多くなった。このディスクキャッシュは磁気ディスク装置を高速アクセスするために必須のサポート項目となっている。

【０００５】上述したライトバック方式のディスクキャッシュを用いる場合には、磁気ディスクコントローラの故障後に於いて、キャッシュメモリに残ったデータを磁気ディスク装置へ書き込む必要がある。従って、単純にキャッシュメモリを持つ磁気ディスクコントローラを２個用意する従来の二重化方式では、磁気ディスク装置に対しての末書き込みデータの書き込みが不可能である。

【０００６】そこで、この不都合を回避するため、キャッシュメモリを磁気ディスクコントローラから切り放し、双方の磁気ディスクコントローラからのアクセスパスを持たせる方式もあるが、このような方式を用いる場合には磁気ディスクコントローラを二重化しない場合にも特殊なキャッシュデバイスを接続する必要があり、システムのハードウェア構造が複雑となり高価となる問題があった。

【０００７】

【発明が解決しようとする課題】上記したように、ライトバック方式のディスクキャッシュを用いる場合には、磁気ディスクコントローラ故障後にキャッシュメモリに残ったデータを磁気ディスク装置へ書き込む必要があり、従って、単純にキャッシュメモリを持つ磁気ディスクコントローラを２個用意する従来の二重化方式では、磁気ディスク装置への末書き込みデータの書き込みが不可能である。そこで、キャッシュメモリを磁気ディスクコントローラから切り放し、双方の磁気ディスクコントローラからのアクセスパスを持たせる方式も考えられたが、この方式は、磁気ディスクコントローラを二重化しない場合にも特殊なキャッシュデバイスを接続する必要があり、システムのハードウェア構造が複雑となり高価になるという問題があった。

【０００８】本発明は上記実情に鑑みなされたもので、

3

通常時に片系磁気ディスクコントローラのキャッシュデータを他系に複製しておくことにより、ライトバック方式のディスクキャッシュを有する磁気ディスク装置の高信頼性を容易に実現する環境を持った磁気ディスクコントローラの二重化方式を提供することを目的とする。

【0009】

【課題を解決するための手段】この発明は、磁気ディスク装置を複数の磁気ディスクコントローラで共有し、ホスト装置によって生成されるデータを上記磁気ディスクコントローラが受信し、ホスト装置に対してその時点で処理終了の旨のステータスを返し、その後に磁気ディスク装置に対するデータの書き込みを行うディスクキャッシュシステムに於いて、双方の磁気ディスクコントローラ間で通信を行うパスを、共有する磁気ディスク装置に対するアクセスパスとして備え、このパスを介して一方の磁気ディスクコントローラが持つキャッシュデータを他方のキャッシュに複製し、双方の磁気ディスクコントローラにてデータの認知を行ない、ホスト装置によって指示される磁気ディスクコントローラ経由で磁気ディスク装置に対してデータの書き込みを行うことを特徴とする。又、稼働状態にある磁気ディスクコントローラは、待機状態にある磁気ディスクコントローラに対し常にキャッシュデータを複製して送付し、稼働状態にある磁気ディスクコントローラの故障時に、待機状態にある磁気ディスクコントローラを介してキャッシュデータを磁気ディスク装置に対し書き込むことを特徴とする。

【0010】

【作用】この発明は、通常時にホスト装置からのコマンドに応答する磁気ディスクコントローラ系から、その磁気ディスクコントローラ系が故障した時に稼働する、通常スタンバイ状態にある磁気ディスクコントローラ系へ、キャッシュデータの内容とデータの格納先アドレス等の制御情報を複写するデータ転送経路と処理手順を設け、通常時は、両系の磁気ディスクコントローラのキャッシュに同一の内容が存在するようにして稼働させ、片系磁気ディスクコントローラの故障時には、残る別系の磁気ディスクコントローラからキャッシュデータを磁気ディスクへ書き込む方式を採ることにより、ライトバック方式のディスクキャッシュを有する磁気ディスクコントローラの二重化を実現する。

【0011】そこで上記した処理を実現するため、各磁気ディスクコントローラにＣＰＵを内蔵し、このＣＰＵの制御により、図２に示すように、他方の磁気ディスクコントローラに対して書き込みアドレスを含むコマンドを発行し、ホスト装置から得られるデータを自系のキャッシュへ書き込むと同時に上記アクセスパスを介して他方の磁気ディスクコントローラに設けた他系のキャッシュへも書き込み、ホスト装置に対してステイタスを返した後、他方の磁気ディスクコントローラからコマンドを受信し、他方の磁気ディスクコントローラが磁気ディスク装置に対し書き込んだデータのディレクトリ情報を得、自系のディレクトリィ更新に備える。

4

【0012】このことにより、ライトバック方式のディスクキャッシュを用いない従来の磁気ディスクコントローラの二重化と同様、さほど信頼性を必要としない場合には磁気ディスクコントローラを二重化せず、高信頼性が必要な場合には磁気ディスクコントローラのみを追加することにより、容易に高信頼システムを構築できる。

【0013】

【実施例】以下、図面を使用してこの発明の一実施例について説明する。図１はこの発明の一実施例を示すブロック図である。図に於いて、符号１は図示しないホスト装置と接続するために用いられるホストパスであり、ホスト装置はこのパス１を介して磁気ディスクコントローラ内のディスクキャッシュ２との間のデータ転送を行う。

【0014】符号２は磁気ディスクコントローラ内のディスクキャッシュであり、ホストパス１、及び後述するディスクパス３とのデータ転送を行う。符号３はディスクパスであり、磁気ディスク装置５へのコマンド・データの送受とともに、例えばＳＣＳＩパス経由での磁気ディスクコントローラ間通信が行える。

【0015】符号４は磁気ディスクコントローラ内のマイクロプロセッサ（ＣＰＵ）であり、磁気ディスクコントローラの全体の処理制御を行う。符号５は磁気ディスク装置であり、ディスクパス３を介して複数の磁気ディスクコントローラ１０，２０により共有使用される。ここでは、磁気ディスクコントローラ１０をマスタ側の磁気ディスクコントローラ、磁気ディスクコントローラ２０をスレーブ側の磁気ディスクコントローラと称す。

【0016】図２はこの発明の実施例の動作を示すフローチャートである。図２（ａ）はマスタとなる磁気ディスクコントローラ１０内のマイクロプロセッサ４のライトコマンド処理の動作手順を示し、同図（ｂ）はスレーブとなる磁気ディスクコントローラ２０内のマイクロプロセッサ４のライトコマンド処理の動作手順を示している。

【0017】図２に於いて、符号Ｓａは磁気ディスクコントローラ１０が他系の磁気ディスクコントローラ２０へコマンドを発行するステップであり、データの格納先アドレスを含む制御情報を他系の磁気ディスクコントローラ２０に通知し、データ転送の準備を行わせる。

【0018】符号Ｓｂは磁気ディスクコントローラ１０がデータ転送を行うステップであり、ここでホストパス１を介して得られるデータを自系の（磁気ディスクコントローラ１０がもつ）ディスクキャッシュ２へ書き込むと同時に、ディスクパス３を経由して他系の（磁気ディスクコントローラ２０がもつ）ディスクキャッシュ２へ書き込む。

【0019】符号Ｓｃは磁気ディスクコントローラ１０

5

が他系の磁気ディスクコントローラ２０から発せられるコマンドを受け取るステップであり、他系の磁気ディスクコントローラ２０が磁気ディスク装置５へ書き込んだデータのディレクトリ情報を受取り、自系の磁気ディスクコントローラ１０によるディレクトリ更新に備える。

【００２０】符号Ｓｄは磁気ディスクコントローラ２０がディスク書き込みを行うステップであり、通常、スタンバイ状態にある磁気ディスクコントローラ２０が負荷分散の意味で処理を受け持つ。符号Ｓｆは磁気ディスクコントローラ２０がディレクトリィの更新要求を発行するステップであり、ディスクキャッシュ２から吐き出されたディレクトリを他系の磁気ディスクコントローラ１０に通知して両系のディスクキャッシュ内容の整合性をとる。

【００２１】以下、この発明の実施例の動作について説明する。第１図に示したように、本発明では磁気ディスクコントローラ間通信が可能なディスクバスを介して２系の磁気ディスクコントローラが接続される。通常はマスタとして示した磁気ディスクコントローラ１０が図示しないホスト装置からのコマンドを受取り、ホスト装置の指示に従った処理を行う。但し、ディスク装置５へのデータ書き込み要求に限り、図２にフローチャートで示す動作を実行する。即ち、自系及びスレイブとなる磁気ディスクコントローラに内蔵のディスクキャッシュにデータを書き込んだ時点で処理を終え、ディスク装置５への書き込み動作自体はスレイブに任せる。

【００２２】スレイブとして示した磁気ディスクコントローラ２０は、マスタとなる磁気ディスクコントローラ１０の故障時のための待機用磁気ディスクコントローラであるが、図２に示したように、待機時にはキャッシュデータを磁気ディスク装置５へデータを書き込む処理を行うライトバックキャッシュコントローラとして動作する。

【００２３】マスタとして示した磁気ディスクコントローラ１０が故障した場合に、スレイブとして示した磁気ディスクコントローラ２０はホスト装置からのコマンドを受付け、磁気ディスク装置５へのコマンド処理全般を行うようになる。

【００２４】このとき、既にマスタとなる磁気ディスクコントローラ１０が処理していた書き込みデータは自系

6

ディスクキャッシュ２内に存在するため、問題なく、その内容も磁気ディスク装置５へ書き込める。

【００２５】一方、スレーブとして示した磁気ディスクコントローラ２０が故障した場合は、マスタとなる磁気ディスクコントローラ１０が磁気ディスク装置５への書き込み処理をスレイブに代わって行う。

【００２６】このとき、必要なデータはやはり自系ディスクキャッシュ２内にあるため、動作の継続は問題なく行える。片系の磁気ディスクコントローラが故障した場合にはその旨の報告はホスト装置に通知される。このとき、片系磁気ディスクコントローラのみでラインバック方式のキャッシュ動作を行わせるか否かはホスト装置の判断に委ねられる。また、磁気ディスクコントローラの二重化を行わない場合にも、図１に示した磁気ディスクコントローラは単体でライトバッファキャッシュ動作が可能であり、信頼性をさほど必要としないシステムでは磁気ディスクコントローラ単体でライトバックキャッシュ動作を行わせ、必要となった場合には磁気ディスクコントローラを追加するのみで信頼性を向上できる。

【００２７】

【発明の効果】以上説明のように、本発明によれば、通常時に片系磁気ディスクコントローラのキャッシュデータを他系磁気ディスクコントローラのディスクキャッシュに複製しておくことにより、容易にコントロールの二重化を実現でき、ライトバック方式のディスクキャッシュを用いない従来の磁気ディスクコントローラの二重化と同様、さほど信頼性を必要としない場合には磁気ディスクコントローラを二重化せず、信頼性が必要な場合にのみ磁気ディスクコントローラを追加することにより、容易に高信頼システムを構築できる。

【図面の簡単な説明】

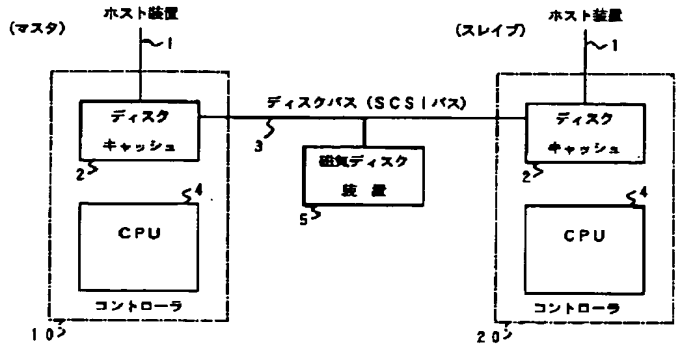
【図１】この発明の実施例の構成を示すブロック図。

【図２】この発明の実施例の動作を示すフローチャート。

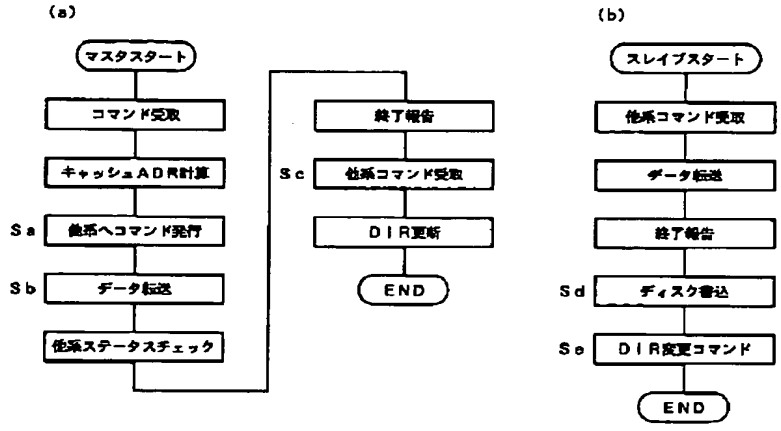
【符号の説明】

１…ホストバス、２…ディスクキャッシュ、３…ディスクバス（ＳＣＳＩバス）、４…マイクロプロセッサ（ＣＰＵ）、５…磁気ディスク装置、１０，２０…磁気ディスクコントローラ。

【図1】


(マスタ)
ホスト装置
1
ディスク
キャッシュ
2
CPU
4
コントローラ
10
ディスクパス（SCSIパス）
3
磁気ディスク
装置
5
(スレイブ)
ホスト装置
1
ディスク
キャッシュ
2
CPU
4
コントローラ
20

【図2】


(a)
マスタスタート
コマンド受取
キャッシュADR計算
Sa
他系へコマンド発行
Sb
データ転送
他系ステータスチェック
終了報告
Sc
他系コマンド受取
DIR更新
END
(b)
スレイブスタート
他系コマンド受取
データ転送
終了報告
Sd
ディスク書込
Se
DIR変更コマンド
END